Artificial Intelligence

Course information

AI Slides 9e, 2024

人工智能讲义, 第 9 版

林作铨 (Lin Zuoquan)

Information and Computation Science Department
Peking University
linzuoquan@pku.edu.cn

Course homepage
www.math.pku.edu.cn/teachers/linzq/ai
course.pku.edu.cn

The chapter-by-chapter list is a syllabus that is subject to lecture-per-week-per as scheduled

Evaluation & homework

Evaluation consists of usual performances (homework) and final exam

- up to 20-30% (graduate-undergraduate) proportion of total evaluation for the homework (without midterm test)
- only final (closed) exam in the last class

Homework (in the separate file) is required to submit to the TA

- Write in English or Chinese by LaTeX (suggestion)
- Submit last week's homework before the next class by the pdf file online in `course.pku` (no record of late delivery, except for the first week)

Q&A platform at `piazza.com`

— The TA will announce the information

Office time: see the course homepage

— Send an email with the domain name `pku` to ask for assistance in person

References

Stuart Russell and Peter Norvig
Artificial Intelligence: A Modern Approach 4e (AIMA), Pearson, 2020
人工智能: 现代方法（第4版），人民邮电出版社，2022
The book website: `aima.cs.berkeley.edu`
including codes `github.com/aimacode`
and exercises `aimacode.github.io/aima-exercises`

Courtesy some sources (slides and figures) from the websites (without cited in each slide due to typeset space)

AI Slides: more and the newest materials than AIMA
More references are included in the slides which would be required for reading as the course progresses, and it is encouraged to look for supplemental materials from other books and papers to expand knowledge

Overview

1. **Introduction**
2. **Search Algorithms I**
3. **Search Algorithms II**
4. **Constraint Satisfaction Problems**
5. **Logical Agents**
6. **Automated Reasoning**
7. **Automated Planning**

Breadth: covering all major fields of AI (core requirement)
Depth: going to the front of the fields (including all the slides with *
for self-study, out of the course time)

Overview

8. **Uncertain Knowledge and Reasoning**
9. **Making Decisions**
10. **Machine Learning**
11. **Natural Language Understanding**
12. **Robotics**
13. **Knowledge Representation***
14. **AI Philosophy**†

†: maybe no teaching
+: optional for undergraduate but requirement for graduate students
#: requirement but may be self-taught due to limited time
*: no requirement for the exam, learned as extended knowledge
— in detail each lecture

Introduction

1

1 Introduction

1.1 AI

- Brief history
- The state of the art

1.2 Intelligent agent

- Agent programs
- Agent structures
- Multi-agent systems^{+}

$^{+}$: optional for undergraduate but requirement for graduate students

AI

What is AI??

What is Intelligence?

Can a machine think? (behave like a thinking person)

Thinking is some process that people engage in every day

Intelligence is an intuitive concept

General intelligence is a general mental capability

There is no precise definition of intelligence or thinking

Artificial Intelligence (AI) attempts to understand intelligence entities, strives to building intelligent agents that perceive and act in an environment, and makes computers smarter in human-level intelligence

- understanding the principle of human intelligence
- making intelligent machines to replace human works

Does machine intelligence differ from human one??

Machine Intelligence does not imply thinking in a human-like way

Intelligence and computation

Computation (or computable by algorithm) is an intuitive concept
– an explicit effective set of instructions to find the answers to any of a given class of problems in finite steps
can be precisely defined by the computational models (computability)
Turing machine, recursive functions, automata etc.
all these computational models are identical
i.e., the class of problems computable by the algorithm is identical to the class of problems solved by the computational models

Computation is typically carried out by an electronic digital computer, but might also be carried out by a person or by a mechanical device of some sort (machine)
– simulating intelligence by computation

It failed to precisely define intelligence as something like computation by some mathematical models

AI vs. brain

Big puzzle: brain $\rightarrow$ mind (conscious, thinking, understanding) $\rightarrow$ intelligence

The brain is an existence reference of intelligent machines to imitate

E.g., birds were a reference for heavier-than-air flight
- shouldn't just copy it, like a kite and earlier airplane
- airplanes were inspired by birds
- they use the same basic principles for flight
  aerodynamics and compressible fluid dynamics
- but airplanes don't flap wings and have feathers

AI needs to understand the **principle** of intelligence

What is the equivalent of aerodynamics for understanding intelligence??

Cognition and recognition

Roughly, intelligence is regarded as two levels
Recognition – a process of interpretation of perception and sensory information, e.g., hearing, vision, feeling

Cognition – a mental process of acquiring knowledge and understanding through thought, experience, sense perception etc.

– – knowledge through thought: thinking even without experience or sense perceptions, e.g., concept formation

(cognition does not necessarily depend on sense perception; memory through sense perceptions can aid in cognition)

– – knowledge through experience: your own or someone else

– – knowledge through senses: perceptions can lead to thought processes and acquisition of knowledge

They are closely related in general

Views of AI

Weak AI: a special purpose computer system can solve a problem in some respect of human-level intelligence
– succeeded from the beginning of AI

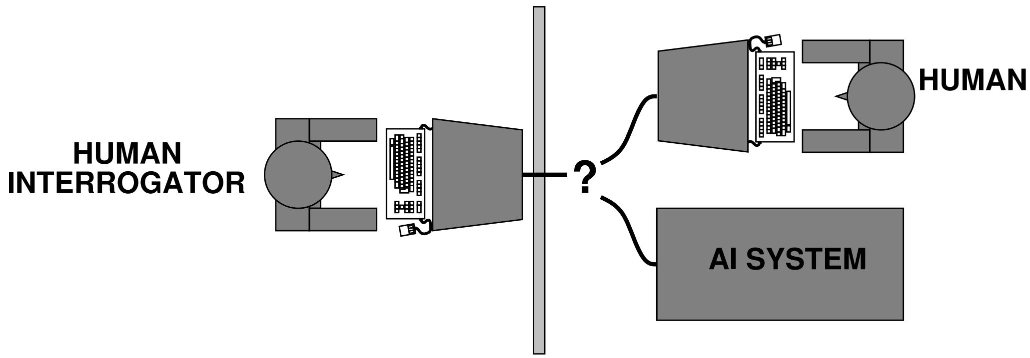
Strong AI: a general-purpose computer system can solve a class of problems in almost all respects of human-level intelligence
– AGI (Artificial General Intelligence)
– not succeeded yet

Views of AI fall into four categories

Thinking humanly	Thinking rationally
Acting humanly	Acting rationally

Acting humanly: The Turing test

- Can a machine think??
- Operational test for intelligent behavior: Imitation Game



- Predicted that by 2000, a machine might have a **30**% chance of fooling a lay person for **5** minutes
- Anticipated all major arguments against AI in following decades
- Suggested major components of AI: knowledge, reasoning, language understanding, learning, etc.

Turing test

The following interaction from Turing's paper

Q: *Please write me a sonnet on the topic of the Forth Bridge.*
A: *Count me out on this one. I never could write poetry.*
Q: *Add 34957 to 70764.*
A: *(Pause about 30 seconds and then give the answer as) 105621.*

Given the fact that you can **fool** some of the people all the time
it is not clear how rigorous this particular standard is

Note: language plays a special role in human behavior, not seen in other animals

– much of how we deal with new situations involves using what we have read or been told earlier using language

Reading: *Turing A, Computing machinery and intelligence, 1950*

Some Turing test programs

- ELIZA, MegaHAL, TIPS, A.L.I.C.E etc.
- Chatbots: MGONZ, NATACHATA, CyberLover etc. (chatbots.org)
- The Chatterbox challenge
- The Loebner Prize for Turing-test-like competition from 1991-2020, no winner

Related tests

- Microsoft Windows Cortana, Apple Siri, Google Assistant, IBM Waston, Amazon Alexa, Facebook Messenger etc. – cannot win
- OpenAI ChatGPT/GPT-4 – not yet

Play "Human or Not" yourself

https://www.humanornot.ai/

Challenge: The Turing test is not reproducible or amenable to mathematical analysis

Thinking humanly: Cognitive Science

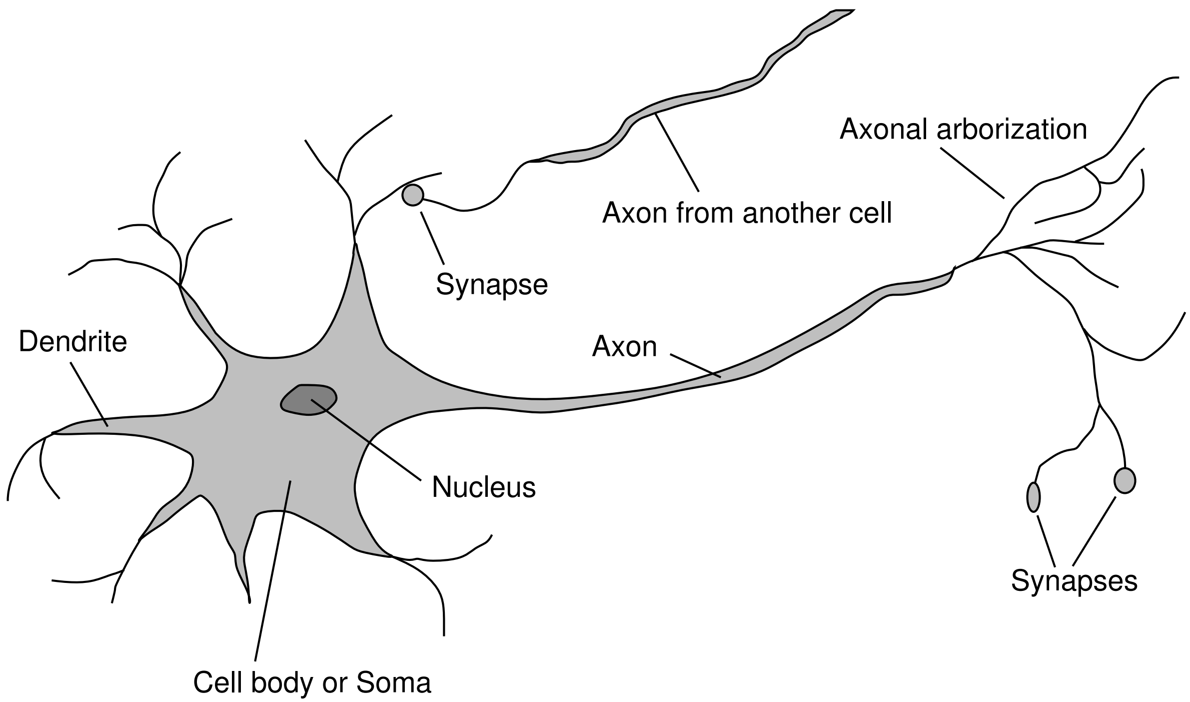
1960s "cognitive revolution": information-processing psychology replaced prevailing orthodoxy of behaviorism

Scientific theories of internal activities of the brain
– What level of abstraction? "Knowledge" or "circuits"
1) Predicting and testing behavior of human subjects (top-down)
2) Direct identification from neurological data (bottom-up)
Both approaches (roughly, Cognitive Science and Cognitive Neuroscience) are now distinct from AI

Cognitive Science and AI share one principal direction

Brains

10^{11} neurons of > 20 types, 10^{14} synapses, 1ms–10ms cycle time
Signals are noisy "spike trains" of electrical potential



Artificial neural networks

Artificial neural networks (ANN/NN): artificial neurons mimic the way biological brain with clusters of biological neurons connected by axons

– an oversimplification of real neurons, but its purpose is to develop an understanding of what networks of simple units can do

The neural network's approach is called **connectionism** in AI

Resurgence under the name deep learning, distinct from the brains and cognitive science

Artificial brain projects

Artificial Brain: direct human brain emulation using artificial neural networks on a high-performance computing engine

- IBM Blue Brain project (grant from Pentagon, 2008)
  Google etc.
- BRAIN Initiative (US, 2013)
  The Human Brain Project (Europ, Japan)
- China Brain Project (China, proposal from 2018)

Challenge: Artificial neural networks and artificial brains are simpler to create general intelligent actions directly without the principle of intelligence

Thinking rationally: Logic

Aristotle: what are correct arguments/thought processes?

Originally, logic is study of thought, or intelligence
but mathematical logic by symbolic method intended to study inferences in mathematics

Various forms of logic:
notation and rules of derivation for thoughts
may or may not have proceeded to the idea of mechanization

The direct line through philosophy, mathematics and logic to AI
so-called logicist in math

The logical approach is called **symbolism** in AI

Knowledge and common sense

Knowledge is the power of intelligence, with especially common sense
– knowing, and solving problems by using knowledge

What is common sense??
How is having common sense any different from being well-trained on large amounts of data?
Common sense is not explained, but
rely on our routines of behavior that we have learned over time
act in situations that are sufficiently unlike the routines we have seen before
Common sense is critical to human-level intelligence and AI

AI would bring logic back to the original goal

Challenge: Not all intelligent behavior is mediated by mathematical logic

Acting rationally: Rational agent

Rational behavior: doing the right thing

right thing: that is expected to maximize goal achievement, given the available information

A rational agent acts to achieve the best (expected) outcome

Viewpoints of rational agents something like engineering

(1) all the skills needed for the Turing test allow an agent to act rationally

(2) logical inference is one but not all of the possible mechanisms for achieving rationality

(3) human behavior is adapted for agent design

No matter symbolism vs. connectionism

Challenge: The rational agent doesn't necessarily involve thinking

Symbolism vs. connectionism

Symbolism & cognition

- Classical AI/Good Old-Fashion AI (GOFAI)/ knowledge-based AI: symbolism toward cognition in AI ⇐ main stream before 2000
- E.g., logic can not deal with hearing
- Foundations of AI, such as problem-solving, knowledge, reasoning, planning, decision, machine learning (except for deep learning), natural language understanding etc.

Connectionism & recognition

- Generative AI/Adaptive AI/data-driven AI: connectionism toward recognition ⇐ main stream after 2000
- E.g., neural networks can not deal with a deduction (100% correct)
- Active in AI, that is deep learning and its applications such as computer vision, speech, natural language processing, etc.

Symbolism vs. connectionism

Two different approaches to AI

- connectionism: current popular in AI, industry and society
- symbolism: breakthrough anytime

Challenging each other

- cycle between "spring" and "winter"

Keep constant temperature: good manner in the study of AI

Neurosymbolic methods: integration of symbolism and connectionism

- All share one principal direction
- The available theories do not explain anything resembling human-level general intelligence

What are the principles of intelligence??

— What are the principles of human intelligence?
— What is machine intelligence?

In principle, you have to learn all AI

Example: LEAN

An interactive proof assistant for writing and checking mathematical proofs $\Leftarrow$ symbolism (logic)

- Users write math proofs using the formal language
- LEAN checks the proofs for correctness and consistency
- If the proof is valid, LEAN provides a formal verification of the theorem or construction

• Research: formalize and verify math theorems and constructions (formalizing math)

• Industry: formal verification of mathematical properties and computer software, etc.

• Education: as a tool for teaching formal reasoning, logic, and proof techniques

`leanprover.github.io`

Example: ChatGPT

A conversational assistant (model) for generating human-like text (video/audio) $\Leftarrow$ connectionism (deep learning)

- Users type a query or prompt into the chat interface
- ChatGPT will generate a response
- engaging in conversation (follow-up asking), utilizing the output, refining and iterating, evaluating and applying

- Research: Natural (multi-modalities) language processing
- Industry: E-commerce (customer support), content generation, language translation, healthcare, etc.
- Education: Students can use it for study support (including coursework), research assistance, and writing assistance to enhance their learning and productivity (revolution in education)

`chat.openai.com`

— You can find and use many similar models free of charge

Example: Doing math with LEAN & ChatGPT

- Leveraging ChatGPT for math inquiry

  - Conceptual understanding: seeking explanations, examples, and insights into intricate math concepts and theories

  - Problem-solving assistance: ChatGPT can provide step-by-step solutions, offer problem-solving strategies, and assist mathematicians in tackling challenging math problems

- Combining LEAN and ChatGPT

  - Optimizing mathematical exploration: mathematicians can combine the formal reasoning capabilities of LEAN with the natural language interactions of ChatGPT to optimize their math exploration and research endeavors

  - Holistic problem-solving: the integration of Lean and ChatGPT creates a holistic problem-solving environment, allowing mathematicians to engage in formal reasoning while receiving conversational support for their mathematical inquiries

AI as interdiscipline

Philosophy	logic, methods of reasoning
	mind as physical system
	foundations of learning, language, rationality
Mathematics	formal representation and proof
	probability
(Computer science)	algorithms, computation, (un)decidability, (in)tractability
Psychology	adaptation, perception and motor control
	experimental techniques (psychophysics, etc.)
Linguistics	knowledge representation
	grammar
Neuroscience	physical substrate for mental activity
Control theory	homeostatic systems, stability
	simple optimal agent designs
Economics	decision and operations (e.g., information processing)

AI is a discipline of computer science

Brief history

1943	McCulloch & Pitts: Boolean circuit model of brain
1950	Turing's "Computing Machinery and Intelligence"
1952–69	Early AI (early enthusiasm, great expectations)
1950s	Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
1956	Dartmouth meeting: "Artificial Intelligence" adopted (**AI birth**)
1965	Robinson's complete algorithm for logical reasoning (resolution)
1966–73	AI discovers computational complexity Neural network research almost disappears
1969–79	Early development of knowledge-based systems
1980–	AI becomes an industry
1986–93	Expert systems industry busts: "AI Winter"
1986–	Neural networks return to popularity (deep learning) ALife, GAs, soft computing
1987–	Rapid increase in technical depth of mainstream AI: AI as a science

AI age

1995–	Intelligent agents
2000–	Semantic web and web services
2001	Big data (very large data sets)
2007	Artificial General Intelligence (AGI)
2010–	Smart earth and smart products, Internet of things
	AI embedded in the infrastructure of almost every industry
	(ambient intelligence, human-machine intelligence)
2012	deep learning
2015–	**AI age** coming
2022	Generative AI (say ChatGPT)
-Present	+AI: Internet/internet of things+AI $\rightarrow$ industry+AI
	AI+: AI for science (AI4S) $\rightarrow$ AI for society

AI Index at `aiindex.org` (reports on the state of the art in AI)

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong
- Drive safely along a curving mountain road

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong
- Drive safely along a curving mountain road
- Drive safely along in downtown

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong
- Drive safely along a curving mountain road
- Drive safely along in downtown
- Buy a week's worth of groceries on the web

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong
- Drive safely along a curving mountain road
- Drive safely along in downtown
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at supermarket

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong
- Drive safely along a curving mountain road
- Drive safely along in downtown
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at supermarket
- Win the national championship Chinese chess

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong
- Drive safely along a curving mountain road
- Drive safely along in downtown
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at supermarket
- Win the national championship Chinese chess
- Discover and prove a new mathematical theorem

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong
- Drive safely along a curving mountain road
- Drive safely along in downtown
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at supermarket
- Win the national championship Chinese chess
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong
- Drive safely along a curving mountain road
- Drive safely along in downtown
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at supermarket
- Win the national championship Chinese chess
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology
- Write an intentionally funny story

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong
- Drive safely along a curving mountain road
- Drive safely along in downtown
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at supermarket
- Win the national championship Chinese chess
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology
- Write an intentionally funny story
- Give competent legal advice in a specialized area of law

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong
- Drive safely along a curving mountain road
- Drive safely along in downtown
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at supermarket
- Win the national championship Chinese chess
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology
- Write an intentionally funny story
- Give competent legal advice in a specialized area of law
- Translate spoken English into spoken Chinese in real time

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong
- Drive safely along a curving mountain road
- Drive safely along in downtown
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at supermarket
- Win the national championship Chinese chess
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology
- Write an intentionally funny story
- Give competent legal advice in a specialized area of law
- Translate spoken English into spoken Chinese in real time
- Converse successfully with another person for an hour

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong
- Drive safely along a curving mountain road
- Drive safely along in downtown
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at supermarket
- Win the national championship Chinese chess
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology
- Write an intentionally funny story
- Give competent legal advice in a specialized area of law
- Translate spoken English into spoken Chinese in real time
- Converse successfully with another person for an hour
- Perform a complex surgical operation

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong
- Drive safely along a curving mountain road
- Drive safely along in downtown
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at supermarket
- Win the national championship Chinese chess
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology
- Write an intentionally funny story
- Give competent legal advice in a specialized area of law
- Translate spoken English into spoken Chinese in real time
- Converse successfully with another person for an hour
- Perform a complex surgical operation
- Walk with robot secretary in downtown

The state of the art

Which of the following can be done at present?

- Play a decent game of ping-pong
- Drive safely along a curving mountain road
- Drive safely along in downtown
- Buy a week's worth of groceries on the web
- Buy a week's worth of groceries at supermarket
- Win the national championship Chinese chess
- Discover and prove a new mathematical theorem
- Design and execute a research program in molecular biology
- Write an intentionally funny story
- Give competent legal advice in a specialized area of law
- Translate spoken English into spoken Chinese in real time
- Converse successfully with another person for an hour
- Perform a complex surgical operation
- Walk with robot secretary in downtown

AI for science

Generative AI has revolutionized the way scientists conduct research and make discoveries

- Mathematics: solving complex math problems, proving theorems, generating new conjectures, and replacing classical math methods Major achievement: The use of AI to solve mathematical problems that were previously unsolvable
- Physics: analyzing complex physical systems, optimizing experimental designs, and predicting the behavior of quantum systems
- Chemistry: predicting chemical reactions, designing new molecules, and optimizing drug discovery processes
- Biology: enabling the analysis of large genomic and proteomic datasets, say AlphaFold • Earth sciences: analyzing satellite imagery, modeling climate change, and predicting natural disasters
- Astronomy, etc.

Intelligent agents

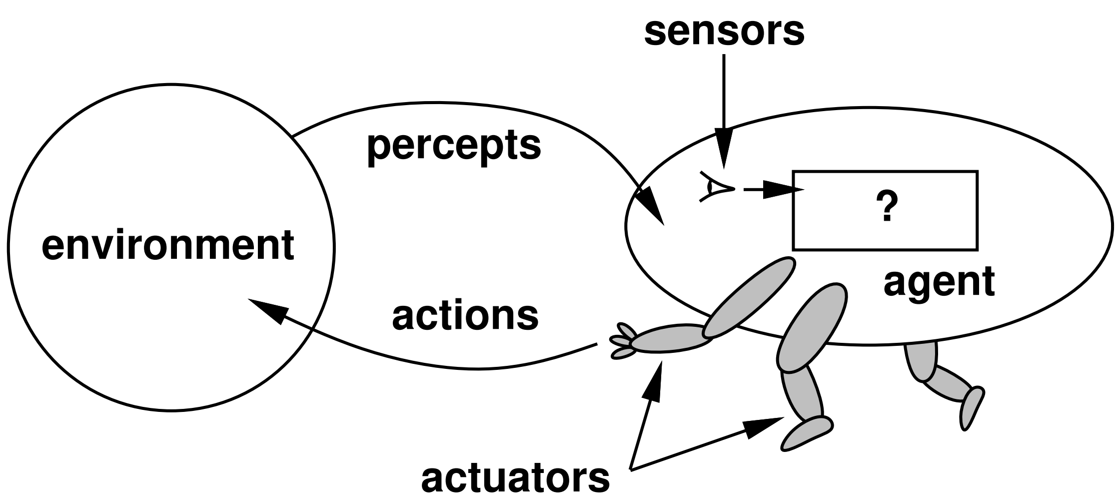
An agent is an entity that perceives and acts in an environment

Agents include

– animal agents
– human agents
– robotic agents (robots)
– software agents (softbots)
– – internet agents
– – – crawler
– – – webbot
– – – email agent
– – – search agent, etc.
– – chatbots
– etc.

Single agent or usually multi-agents (so-called distributed AI)

Agents and environments



An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators

Sensors and actuators#

A sensor measures some aspect of the environment in a form that can be used as input by an agent

– vision, hearing, touch

– active sensing: send out a signal (such as radar or ultrasound) and sense the reflection of this signal off of the environment

⇒ IoT (Internet of Things)

Perception provides agents with information about the world they inhabit by interpreting the response of sensors

Actuators

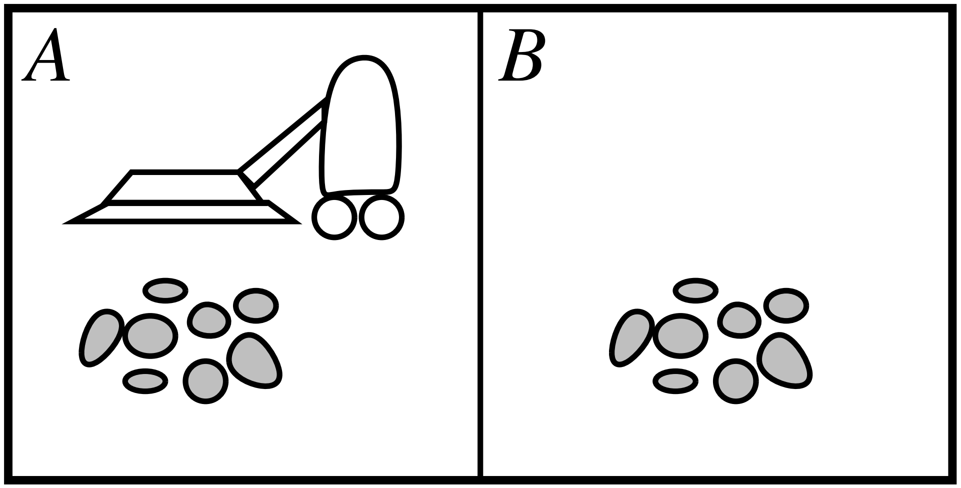
– hands, legs, vocal tract etc.

– robotaxi: those available to a human driver
e.g., accelerator, steering, braking, and so on

#: requirement but may be self-taught due to limited time

Example: Vacuum-cleaner world

Example: Robot cleaner, say, iRobot Roombat



Percepts: location and contents, e.g., $[A, Dirty]$

Actions: $Left$, $Right$, $Suck$, $NoOp$

Agent functions

An agent is completely specified by the agent function: maps from percept histories to actions

$$f : \mathcal{P}^* \rightarrow \mathcal{A}$$

For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance

Find a way to implement the rational agent function concisely

Example: A vacuum-cleaner agent

Percept sequence	Action
$[A, Clean]$	$Right$
$[A, Dirty]$	$Suck$
$[B, Clean]$	$Left$
$[B, Dirty]$	$Suck$
$[A, Clean]$, $[A, Clean]$	$Right$
$[A, Clean]$, $[A, Dirty]$	$Suck$
⋮	⋮

What is the **right** function? Can it be implemented in a program?

Agent programs

The agent program runs on the physical architecture to produce the agent function

$agent = architecture + program$

$program = algorithm + data$

The agent program takes a single percept as input and keeps internal state

```
def SKELETON-AGENT(percept)
    persistent: memory, the agent's memory of the world

    memory ← UPDATE-MEMORY(memory, percept)
    action ← CHOOSE-BEST-ACTION(memory)
    memory ← UPDATE-MEMORY(memory, action)
    return action
```

The algorithm is described by the pseudocode

Other forms of algorithm

```
function SKELETON-AGENT( percept) returns action // output result
    inputs: percept, input from sensors  // may be omitted
    persistent: memory, the agent's memory of the world

    memory ← UPDATE-MEMORY(memory, percept)
    action ← CHOOSE-BEST-ACTION(memory)
    memory ← UPDATE-MEMORY(memory, action)
    return action  // output
```

The algorithm is described by the pseudocode

Other forms of algorithm

A procedure for the skeleton-agent that takes a single percept as input, keeps internal state

Procedure of Skeleton-Agent
Input: *percept*
Output: *action*
memory the agent's memory of the world

1. $memory \leftarrow \text{UPDATE-MEMORY}(memory, percept)$
2. $action \leftarrow \text{CHOOSE-BEST-ACTION}(memory)$
3. $memory \leftarrow \text{UPDATE-MEMORY}(memory, action)$

The algorithm is described by the pseudocode

Other forms of algorithm

Popular forms of pseudocode are written by LaTeX algorithmic packages (e.g., algorithm2e)

```
procedure SKELETON-AGENT(percept)
    memory: the agent's memory of the world
  memory ← UPDATE-MEMORY(memory, percept)
  action ← CHOOSE-BEST-ACTION(memory)
  memory ← UPDATE-MEMORY(memory, action)
  return action
```

Algorithm

An algorithm is an explicit effective set of instructions for a computing procedure

– may be used to find the answers to any of a given class of questions

– can be precisely defined by computational models, e.g., Turing machine, etc.

Analysis of algorithms, independently of the particular implementation and input

– time complexity: speed in seconds

e.g., the worst $T_{worst}(n)$ or the average $T_{avg}(n)$

– space complexity: memory consumption in bytes

Complexity analyzes problems rather than algorithms

Algorithm

O-notation: asymptotic analysis

$T(n)$ is $O(f(n))$ if $T(n) \leq kf(n)$ for some k, for all $n > 0$

$O(n^k)$ for some k: polynomial time/space, called P or "easiness"
Otherwise, exponential time/space, e.g., $O(2^n)$, called NP, hardness
– a class of nondeterministic (Turing machine) polynomial problems
guess a solution and then verify whether the guess is correct in polynomial time
– NP-complete problems (the hardest subclass of NP)
– – co-NP (-complete) is the complement of NP (-complete)
"yes" and "no" answers reversed

Theorem: either all the NP-complete problems are in P or none of them is

Algorithm$^{\#}$

$NP = P$??

if you could try all the guesses at once, or you were very lucky and always guess right the first time

$\#P$: counting problems corresponding to decision problems in NP

decision problems have a yes-or-no answer, counting problems have an integer answer: how many solutions

at least as hard as any NP problems

$PSPACE$: require a polynomial amount of space, even on a non-deterministic machine

it is believed that $PSPACE$-hard problems are worse than NP-complete problems although it could turn out that $NP = PSPACE$, just as $P = NP$

The pseudocode

The pseudocode is a simple language to describe algorithms
– similar to programming languages like C, Java, Lisp or Python
– informal use of mathematical formulas or ordinary English to describe parts

- **Persistent variables** (global and local variables): a variable is given an initial value the first time a function is called and retains that value on all subsequent calls to the function. The agent programs use persistent variables for memory. Variables have lowercase italic names.
- **Function as values**: The value of a variable is allowed to be a function. Functions and procedures have capitalized names
- **Default values for parameters**: "function $F(x, y = 0)$ return a number" means that is an optional argument with default value 0, i.e., the calls $F(3, 0)$ and $F(3)$ are equivalent

The pseudocode

- **Indentation is significant**: the scope of a loop, conditional or statement block
- **Assignment**: "$x \leftarrow value$" means that the right-hand side evaluates to the left-hand side variable
  - **Destructuring assignment**: "$x, y \leftarrow pair$" means that the right-hand side evaluates to a two-element tuple, and the first element is assigned to x and the second to y. Or "for each x, y in $pair$ do"
- **Conditional** if c then $\cdots$ (else) $\cdots$: if the condition c is hold then doing something; otherwise doing something else

The pseudocode

- **Loops**
  - for x in c do: the loop is executed with the variable x bound to successive elements of the collection c
    or for each x in c do
  - for $i = 1$ to n do: the loop is executed with bound to successive integers from 1 to n inclusive
  - while c $(condition)$ do": means the condition is evaluated before each iteration of the loop, and the loop exits if the condition is false
    or loop $c \cdots$
  - repeat $\cdots$ until c: the loop is executed unconditionally the first time, then the condition is evaluated, and the loop exits if the condition is true; otherwise the loop keeps executing (and testing at the end)

The pseudocode

- **Generator**: a function that contains yield is a **generators** that generates a sequence of values, one each time the yield expression is encountered. After yielding, the function continues execution with the next statement, say, "generator $G(x)$ yields number" defines G as a generator function
- **Data types**: data structure
  - **Lists**: $[x, y, z]$, $[first|rest]$
  - **Sets**: $\{x, y, z\}$
  - **Arrays** start at 1: as in usual mathematical notation, not 0 as in Python
- // the explanations can be given as remark

The code*

The algorithms in the pseudocode can be implemented in Python, Java, C/C++, Lisp and Prolog etc.

The code for each topic is divided into four directories:

– `agents`: code defining agent types and programs
– `algorithms`: code for the methods used by the agent programs
– `environments`: code defining environment types, simulations
– `domains`: problem types and instances for input to algorithms

(Often run algorithms on domains rather than agents in environments)

*: no requirement for the exam and learned as extended knowledge

A Python coding*

```python
import environment as en
import domain as do
import action as ac

def skeleton_agent(percept):
    m = []
    m = do.push(percept)
    a = ac.choose()
    #chosen from action class
    m = ac.update()
    return a
```

Note: encourage to code by Python, but programming language and programming methodology are prerequisites which are out of the basic requirement of this course

Example: A vacuum-cleaner agent

```
def TABLE-DRIVEN-AGENT(percept)
    persistent: percepts, a sequence, initially empty
                table, a table, indexed by percept sequences, initially fully specified

    append percept to the end of percepts
    action ← LOOKUP(percepts, table)
    return action
```

Python: A vacuum-cleaner agent*

```python
def Table_Driven_Vacuum_Agent(table):
    """

    A table is provided as a dictionary of all
    {percept_sequence:action} pairs.
    """
    percepts = []

    def agent(percept):
        percepts.append(percept)
        action = table.get(tuple(percepts))
        return action

    return agent
```

Hint: It can be executable Python code

Python: One more*

```
def Table_Driven_Vacuum_Agent():
    """Tabular approach towards vacuum world

    >>> agent = Table_Driven_Vacuum_Agent()
    >>> environment = VacuumEnvironment()
    >>> environment.add_thing(agent)
    >>> environment.run()
    >>> environment.status == {(1,0):'Clean' , (0,0) : 'Clean'}
    True
    """

    table = {((loc_A, 'Clean'),): 'Right',
             ((loc_A, 'Dirty'),): 'Suck',
             ((loc_B, 'Clean'),): 'Left',
             ((loc_B, 'Dirty'),): 'Suck',
             ((loc_A, 'Dirty'), (loc_A, 'Clean')): 'Right',
             ((loc_A, 'Clean'), (loc_B, 'Dirty')): 'Suck',
             ((loc_B, 'Clean'), (loc_A, 'Dirty')): 'Suck',
             ((loc_B, 'Dirty'), (loc_B, 'Clean')): 'Left',
             ((loc_A, 'Dirty'), (loc_A, 'Clean'), (loc_B, 'Dirty')): 'Suck',
             ((loc_B, 'Dirty'), (loc_B, 'Clean'), (loc_A, 'Dirty')): 'Suck'}
    return Agent(Table_Driven_Vacuum_Agent(table))
```

Hint: It can be executable Python code

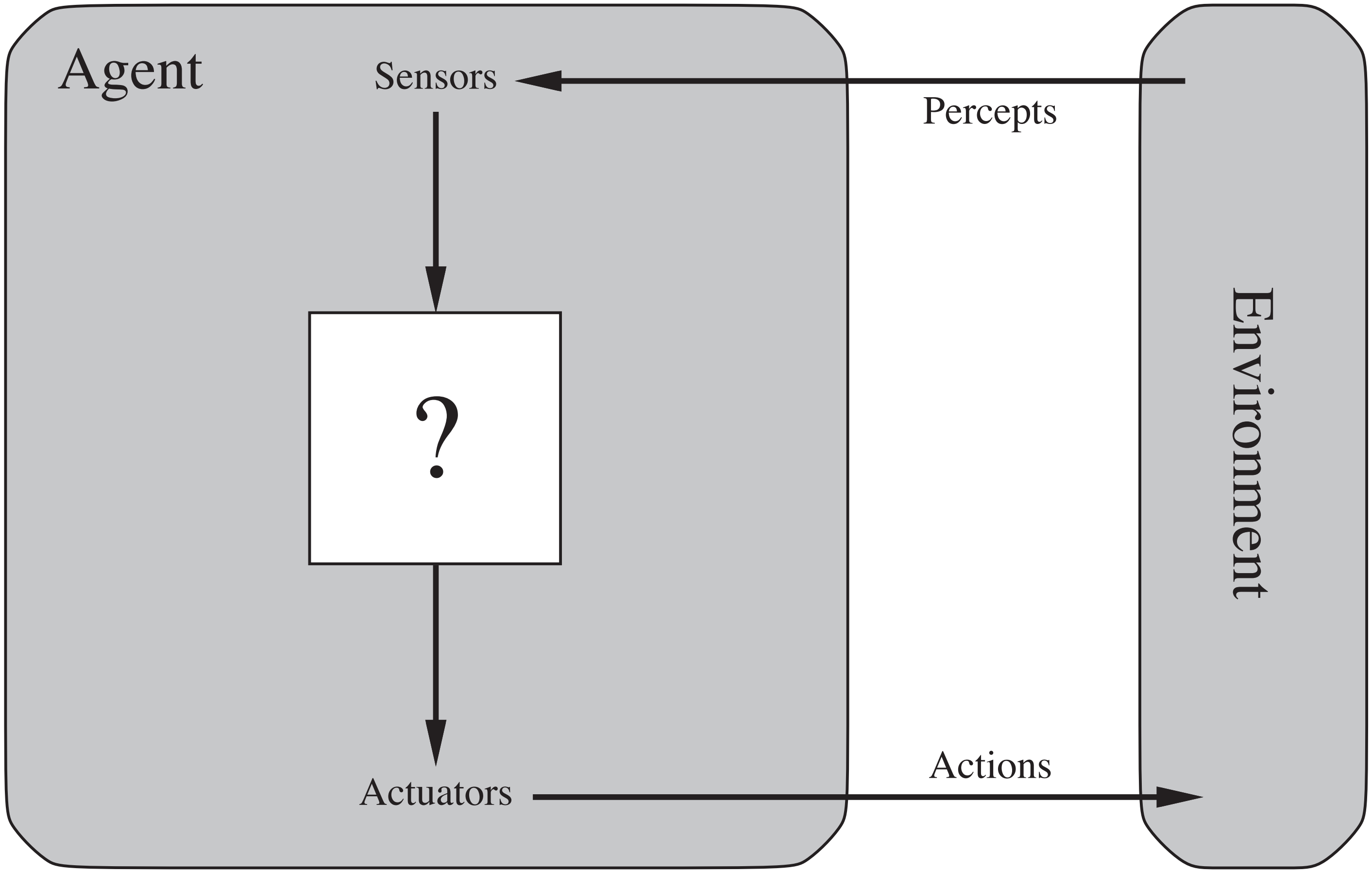
Python: Translating pseudocode to code by ChatGPT*

```python
def Table_Driven_Vacuum_Agent():
    # Define static variables
    percepts = []  # A sequence, initially empty
    table = {     # A table, initially fully specified
        # Percept sequence: Action
        # Add your specific percept sequences and corresponding actions here
        # For example: ('percept1', 'percept2'): 'action1',
    }

    # Append percept to the end of percepts
    percepts.append(percept)

    # Lookup action in the table based on percepts
    action = table.get(tuple(percepts))

    return action
```

Hint: Using ChatGPT or code assistants (say Github Capilot, AlphaCode) to translate the pseudocodes to almost executive codes in popular programming languages

Agent structures

Agents interact with environments through sensors and actuators

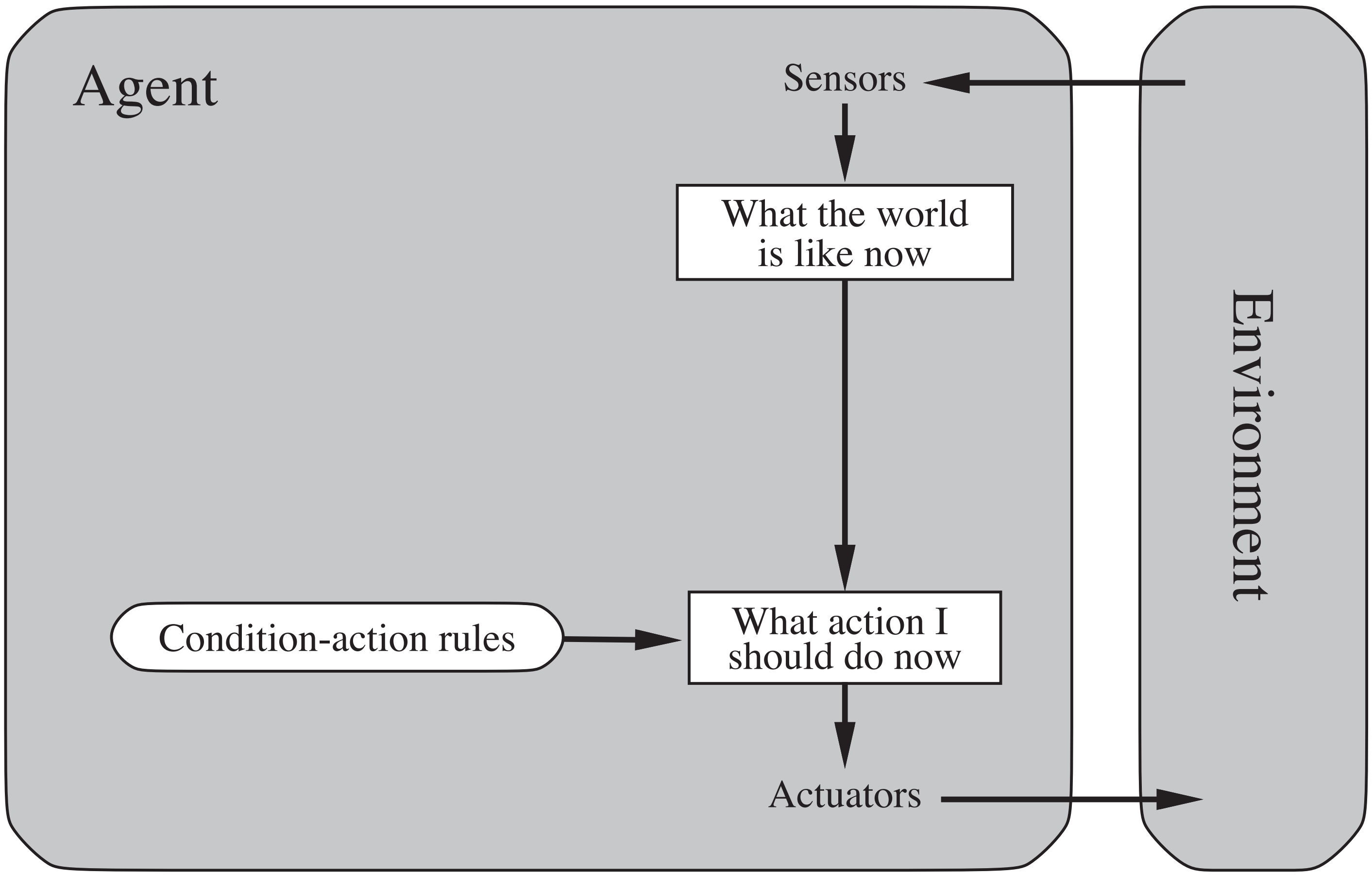


Agent structures

Four basic types in order of increasing generality:
– simple reflex agents
– model-based reflex agents
– goal-based agents
– utility-based agents

All these can be turned into
– learning agents

Simple reflex agents



Example: A vacuum-cleaner agent

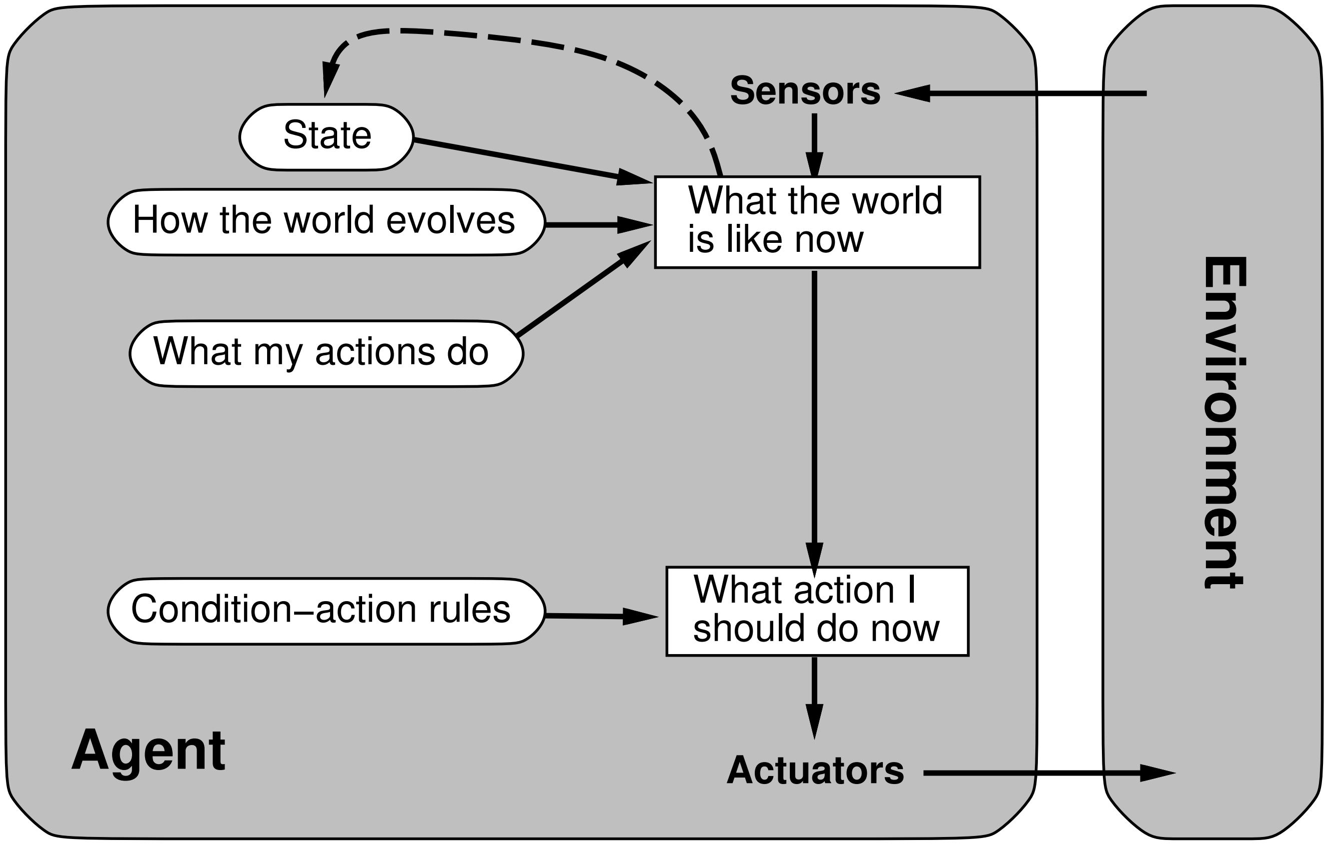
def REFLEX-VACUUM-AGENT([*location*,*status*])

if $status = Dirty$ **then return** *Suck*
else if $location = A$ **then return** *Right*
else if $location = B$ **then return** *Left*

Simple reflex agents

```
def SIMPLE-REFLEX-AGENT( percept)
persistent: rules, a set of condition-action rules

    state ← INTERPRET-INPUT(percept)
    rule ← RULE-MATCH(state, rules)
    action ← rule.ACTION
    return action
```

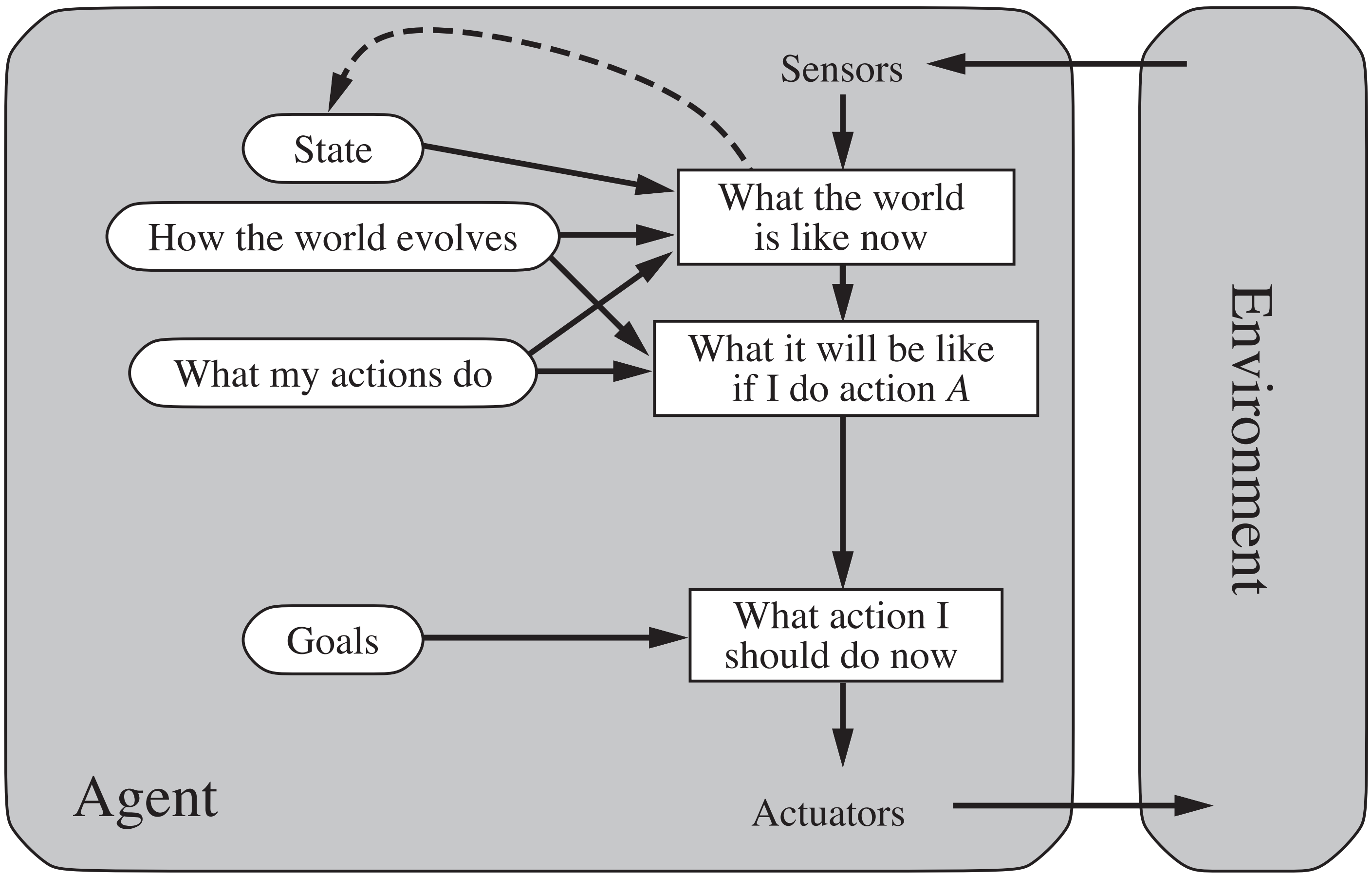
Model-based reflex agents



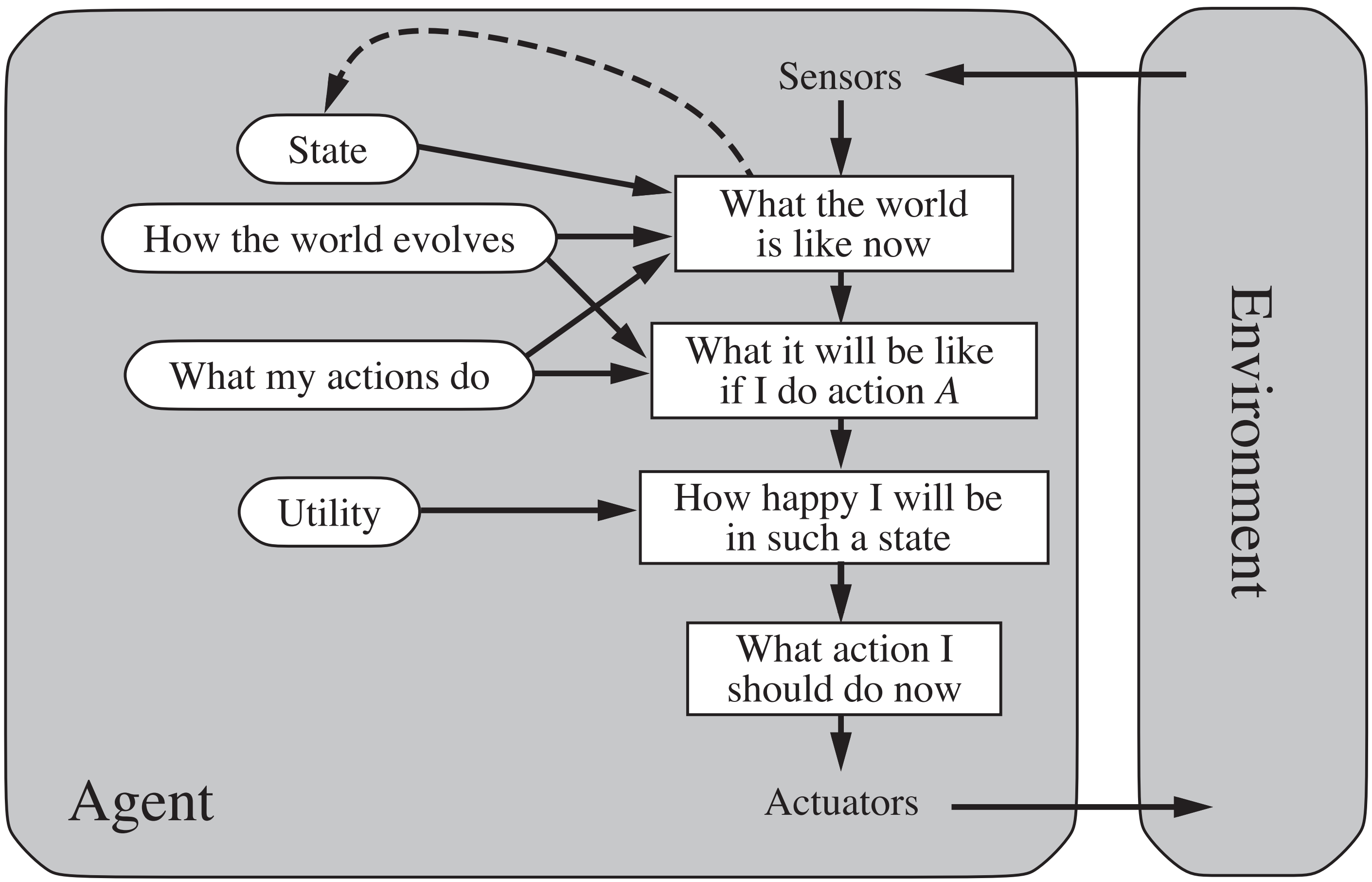
Reflex agents with (internal) state

transition model - how the world works

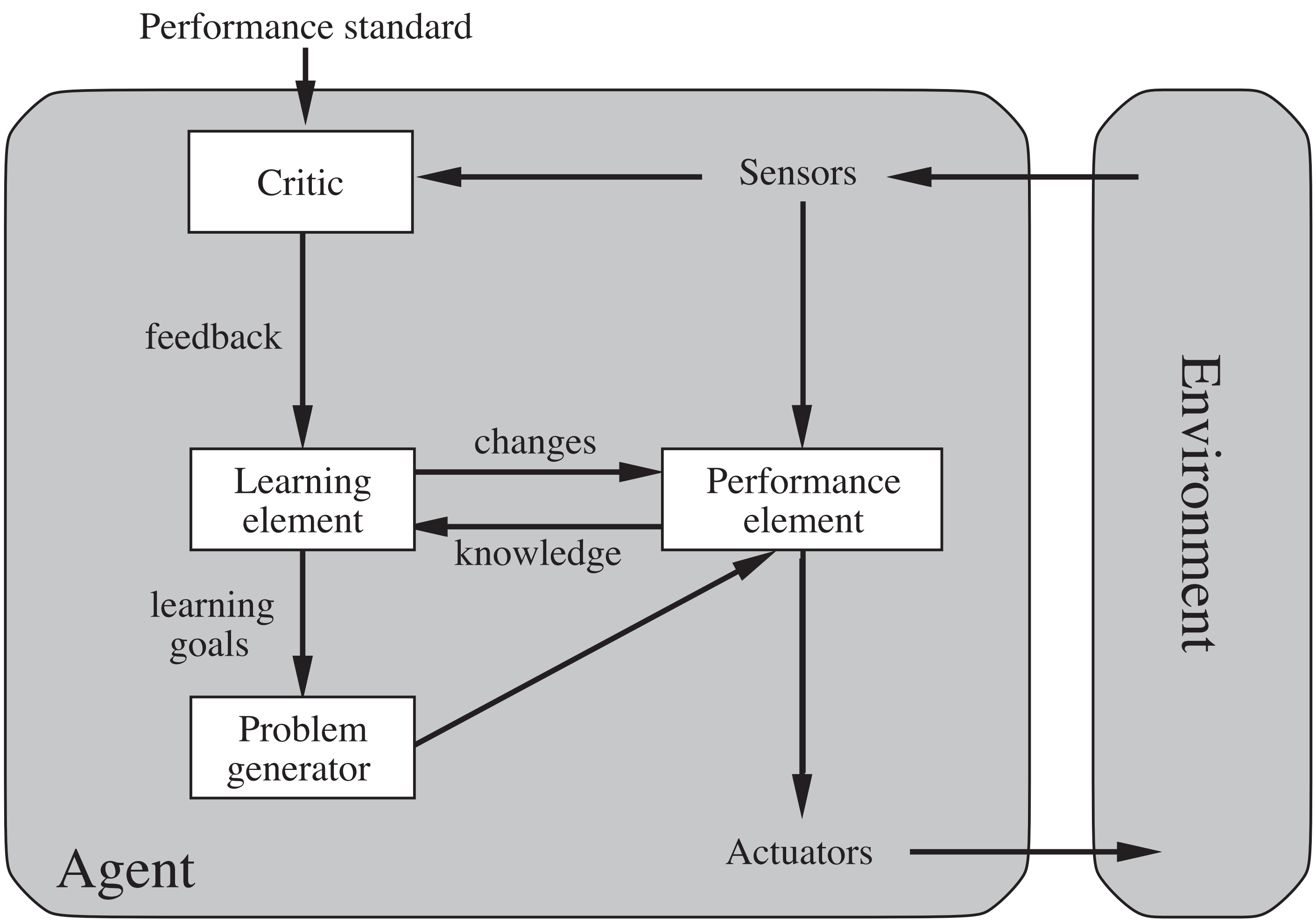
Goal-based agents



Utility-based agents



Learning agents



Multi-agent systems$^+$

MAS (multi-agent systems): a agent interacts with neighbouring agents

a complex task is divided into multiple smaller tasks, each of which is assigned to a distinct agent

Applications

- modeling complex systems
- smart grids
- computer networks
  – cloud computing, social networking, security, routing
- etc.

$^+$: optional for undergraduate but requirement for graduate students

Properties of MAS

Coordination: managing agents to collaboratively reach their goals

. consensus - achieving a global agreement

. controllability - using certain regulations to transmit a state

. synchronization - aligning each agent in time with other agents

. connection - connecting to each other

. formation - organizing in a structure

Communication: communicating among agents

Fault detection and isolation (FDI): a faulty agent may infect other agents that it collaborates with

Task allocation: allocation of tasks to agents considering the associated cost and time

Localization: each agent has a limited view (only its neighbors)

Agent organization

The way agents communicate and connect

- Flat - all agents are regarded as equals
- Hierarchy - agents have tree-like relations
- Holon - agents are organized in multiple groups which are known as holons based on particular features (e.g., heterogeneity), holons are then multiplied layered
- Coalition - agents are temporarily grouped based on their goal
- Team - agents create a team and define a team goal that differs with their own goal
- Matrix - each agent is managed by at least two head agents
- Congregation - agents in a location form a congregation to achieve their requirements that they cannot achieve alone

References†

More reference books

- Readings in AI series (various areas of AI, source papers)
  – E.g., Readings in Knowledge Representation

AI philosophical debate

- Hubert Dreyfus, What Computers Can't do, 1972;
  What Computers Still Can't Do, 1992
  Hubert and Stuart Dreyfus, Mind Over Machine, 1986
- Ray Kurzwell, The Singularity Is Near, 2005
- Nick Bostrom, Superintelligence: Paths, Dangers, Strategies, Oxford University Press, 2014
- etc.

AI fiction†

- Mary Shelley, Frankenstein or Modern Prometheus, 1818
- Samuel Butler, Darwin among the Machines, 1863
- Karel Capek, R.U.R (Rossum's Universal Robots), 1921
- Terry Bisson, They're Made out of Meat, 1990
- too much

AI Movie†

- Future world, 1976
- The terminator, 1984
- The matrix 1-4, 1999-2022
- AI, 2001
- Persons of interest, 2014
- West world, 2016
- too much (appear every year/month)

AI Movie†

Asimov's Three Laws of Robotics

1. A robot must not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where those orders would conflict with the First Law.
3. A robot must protect its own existence, except where such protection would conflict with the First or Second Law.

- Three Laws were clear, direct, and logical. Asimov's stories, on the other hand, told how easily they could fail
- The contradictions in Asimov's laws encouraged others to propose new rules
- Any set of rules will always have conflicts and grey areas

AI News

- You can find AI almost everyday in news, media and everyday life
- Check it, from now on

- At the time of **AI**, work intelligently and enjoy